Rotational Motion Compensated Prediction in HEVC Based Omnidirectional Video Coding

Bharath Vishwanath, Kenneth Rose
University of California, Santa Barbara
Santa Barbara, CA, 93106
{bharathv,rose}@ece.ucsb.edu

Yuwen He, Yan Ye
InterDigital Communications, Inc.
San Diego, CA, 92121
{Yuwen.He,Yan.Ye}@Interdigital.com

***Abstract*—Spherical video is becoming prevalent in virtual and augmented reality applications. With the increased field of view, spherical video needs enormous amounts of data, obviously demanding efficient compression. Existing approaches simply project the spherical content onto a plane to facilitate the use of standard video coders. Earlier work at UCSB was motivated by the realization that existing approaches are suboptimal due to warping introduced by the projection, yielding complex non-linear motion that is not captured by the simple translational motion model employed in standard coders. Moreover, motion vectors in the projected domain do not offer a physically meaningful model. The proposed remedy was to capture the motion directly on the sphere with a rotational motion model, in terms of sphere rotations along geodesics. The rotational motion model preserves the shape and size of objects on the sphere. This paper implements and tests the main ideas from the previous work [1] in the context of a full-fledged, unconstrained coder including, in particular, bi-prediction, multiple reference frames and motion vector refinement. Experimental results provide evidence for considerable gains over HEVC.**

I. INTRODUCTION

Omnidirectional or 360° video is emerging as a new means to offer an immersive visual experience by capturing the video on the sphere, and enabling users to view it in any desired direction. With the increased field of view and hence data rates, spherical video represents a significant compression challenge. Since standard video coders cannot handle spherical content, the video on the sphere is projected onto a plane (or planes) via a projection format, such as equirectangular (ERP), cubemap, octahedron, or icosahedron [2]. Note that, for a given geometry, uniform sampling on the plane induces non-uniform sampling on the sphere, whose density varies with location.

A central compression component of modern video codecs, such as H.264 [3] and HEVC [4], is motion compensated prediction, often referred to as inter-prediction, which is tasked with exploiting temporal redundancy. Most video codecs only allow block-based translational motion for motion compensated prediction, which is incompatible with the complex non-linear motion observed in projected spherical video. Recently an affine motion model was proposed in [5] to effectively capture complex motion in 2D video. However, in 360° video, the amount of warping introduced by the projection varies for different regions of the sphere due to varying sampling density, leading to complex non-linear motion in the projected geometry, which render even the affine motion model ineffective. Both the translational motion model and its affine extension do not preserve the shape and size of objects on the sphere. A relevant recent contribution by Li et al., proposed an interesting 3D translational motion model for the cubemap projection [6]. In this approach, the centers of the current coding block and the reference block are mapped to the sphere and the 3D displacement between these vectors is calculated. The remaining pixels in the current coding block are also mapped to the sphere and then translated by the 3D displacement vector obtained for the block center. These translated vectors are not guaranteed to be on the sphere and thus need to be reprojected to it. Due to this final projection, object shape and size are not preserved, and some distortion is introduced.

A recent work at UCSB addressed this challenge by directly performing motion compensation on the sphere [1]. Specifically, motion was defined as the rotation of the block of pixels on the sphere along geodesics. Rotation, being a unitary transformation, preserves the shape and size of the object on the sphere. Complementary to the rotation motion model, a location invariant motion search pattern was introduced. Rotational motion compensation in conjunction with the location-invariant search grid on the sphere makes the proposed motion model agnostic of the geometry. The objective of that work was to provide a proof of concept for the basic ideas, and the focus was on uni-prediction with single reference frame. The encoder was restricted from refining the best integer motion vector to sub-pixel precision. It was shown that rotational motion compensation gives substantial gains over both HEVC and the 3D translational motion model proposed in [6]. In this paper, we implement the framework of [1] to a full-fledged practical codec, emphasizing the efficacy of the rotational motion model in bi-directional prediction with multiple reference frames. Supporting bi-directional prediction enables us to work with random access and low delay-B profiles. The encoder is further extended to hierarchically refine the integer motion vector to $\frac{1}{4}$ pixel precision, which is consistent with standard HEVC. The substantial gains observed in the experiments validate the utility of the proposed motion model.

The rest of the paper is organized as follows. In section II, we provide an overview of the rotational motion model proposed in [1]. Section III covers the implementation details of the rotation motion model for bi-predictive motion compensation with multiple reference frames and motion vector refinement. Experimental results are provided in Section IV followed by conclusions in Section V.

II. Rotational Motion Model

This section summarizes the concepts proposed in [1] At the heart of the approach is the idea to perform motion compensation directly on the sphere. Let us consider a block of pixels in the projected domain (say ERP) for which we need to derive the prediction signal. An example of such a block in the ERP domain is illustrated in Fig. 1(a). The block of pixels is mapped onto sphere as shown in Fig. 1(b). Let $\mathbf{v}$, the vector on the sphere corresponding to the center of the current prediction block, be motion compensated to a point $\mathbf{v}'$. $\mathbf{v}$ is rotated to $\mathbf{v}'$ along the geodesic connecting them, wherein geodesic is the shortest path between two points on the sphere. Geodesic rotation is shown in Fig. 1(c). Rotation is defined by an axis and angle of rotation. The axis of rotation $\mathbf{k}$ is the vector perpendicular to the plane defined by the origin, $\mathbf{v}$ and $\mathbf{v}'$ and is obtained by taking the cross product of vectors $\mathbf{v}$ and $\mathbf{v}'$, i.e,

$$\mathbf{k} = \frac{\mathbf{v} \times \mathbf{v}'}{|\mathbf{v} \times \mathbf{v}'|}. \tag{1}$$

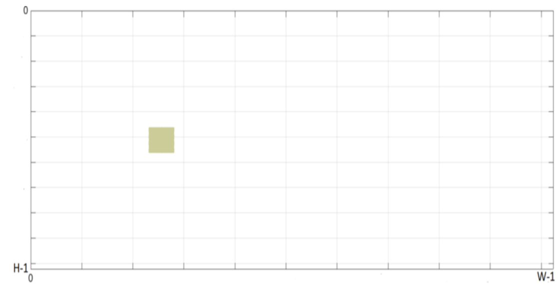
The angle of rotation is given by,

$$\alpha = \cos^{-1}(\mathbf{v} \cdot \mathbf{v}'). \tag{2}$$

Given this axis and angle, all the points in the current block are rotated with the same rotation operation via the Rodrigues' rotation formula [7]. This formula gives an efficient method for rotating a vector $\mathbf{v}$ in 3D space about an axis defined by unit vector $\mathbf{k}$, by an angle α. Let (x, y, z) and (u, v, w) be the coordinates of the vectors $\mathbf{v}$ and $\mathbf{k}$ respectively. The coordinates of the rotated vector $\mathbf{v}'$ will be:

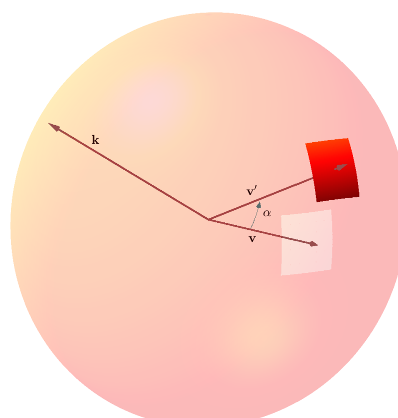
$$\begin{aligned} x' &= u(\mathbf{k} \cdot \mathbf{v})(1 - \cos\alpha) + x\cos\alpha + (-wy + vz)\sin\alpha, \\ y' &= v(\mathbf{k} \cdot \mathbf{v})(1 - \cos\alpha) + y\cos\alpha + (wx - uz)\sin\alpha, \\ z' &= w(\mathbf{k} \cdot \mathbf{v})(1 - \cos\alpha) + z\cos\alpha + (-vx + uy)\sin\alpha \end{aligned} \tag{3}$$

Unlike prior approaches, rotation ensures that all the points that are rotated stay on the surface of the sphere. Moreover, the proposed motion model preserves the shape and size of the object on the sphere. After rotation, the rotated block is mapped to the reference frame. An illustration of rotated block mapped back to ERP domain is shown in Fig. 1(d). Since the projected location might not be on the sampling grid of the reference frame, interpolation is performed in the reference frame to get the pixel value at the projected coordinate. The proposed motion compensation technique is summarized in Algorithm 1.

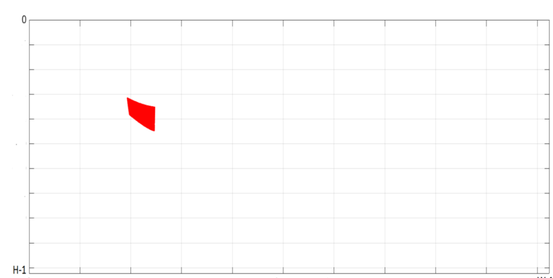


(a) A block in current ERP frame.

(b) Block mapped to sphere



(c) Geodesic rotation of block on sphere



(d) Rotated block mapped to reference ERP frame

Fig. 1. Illustration of various steps in geodesic motion compensation technique

Algorithm 1 Proposed motion compensation technique

1: Map the block of pixels in the current coding unit on to the sphere.
2: Define a location invariant search grid around $\mathbf{v}$, to get $\{\mathbf{v}'\}$
3: Define a rotation operation which rotates $\mathbf{v}$ to $\mathbf{v}'$ along the geodesic from $\mathbf{v}$ to $\mathbf{v}'$.
4: Rotate all the pixels in the block with the rotation operation defined in Step 3.
5: Map the rotated coordinates on the sphere to the reference frame in projected geometry.
6: Perform interpolation in the reference frame to get the required prediction.

III. Generalizing Geodesic Motion Compensation

A. Bi-predictive motion compensation

Modern video coders use bi-prediction for better exploiting the temporal redundancy, wherein the prediction signal is derived by averaging two motion compensated prediction signals from possibly two different frames. We perform an iterative bi-prediction search similar to HEVC, wherein one prediction signal is held fixed and a search is performed to get the best second prediction signal. As mentioned earlier, the translational motion model does not preserve shape and size of the object on the sphere, introducing deformations that vary with the motion vector. Since translation in the projected domain would have introduced different amounts of deformations for the uni-directional prediction signals, simply averaging them would not be appropriate. However, with the rotational motion model, the size and shape of the object is preserved for each prediction signal, irrespective of the motion vector. Thus, averaging the undeformed prediction signals derived from rotations on the sphere leads to effective prediction.

B. Motion Vector Refinement

In [1], the encoder was restricted from performing subpixel motion vector refinement. We note in passing that [1] also included an isotropic search grid which is not implemented here (for now) in order to maintain direct correspondence with the hierarchical motion search employed by HEVC. Instead, we define a new search pattern that fits with the hierarchical motion vector refinement. We treat $\mathbf{v}$, the vector on the sphere corresponding to the center of the current prediction unit, as if it were the vector on the sphere corresponding to zero yaw and pitch. An integer motion vector (m, n) then defines the rotation of $\mathbf{v}$ to a new point $\mathbf{v}'$ whose spherical coordinates (ϕ', θ') are given by:

$$\phi' = m\Delta\phi, \qquad \theta' = n\Delta\theta \tag{4}$$

where, $\Delta\phi$ and $\Delta\theta$ are predefined step sizes. Let H denote the height of the ERP frame, then $\Delta\theta$ is chosen to be $\frac{\pi}{H}$ as it corresponds to the change in the pitch (elevation) when we move by a single integer pixel in vertical direction. Similarly, $\Delta\phi$ is chosen to be $\frac{2\pi}{W}$, where W denotes the width of the ERP frame. For each successive stage of motion vector refinement, the step sizes are correspondingly halved, leading to a direct correspondence with the hierarchical motion vector refinement in HEVC.

C. Multiple reference frames

Video coders allow deriving prediction signal from multiple references. As the temporal distance of the reference frame increases, the resulting motion vector is likely to be of higher magnitude. The resulting prediction signal thus derived by the translational motion model will be highly deformed, further compromising the prediction quality. However, rotation does not deform objects and is hence immune to such suboptimalities, thereby enabling effective multi-reference prediction. In our implementation, reference frames in a reference picture list are searched, similar to the procedure in HEVC, to obtain the best prediction signal.

IV. Experimental Results

The proposed motion model was implemented in HM-16.15 [8]. Geometry mappings were performed using 360Lib-3.0 [9]. The JVET test methodology for 360 degree video described in [10] was used in the experiments for ERP with random access setting. In order to save in computation time, we encoded all the ERP sequences at 2K resolution. We encoded full length of six video sequences over the four QP values of 22, 27, 32 and 37. We used DCT-IF filter [11] at $\frac{1}{16}^{th}$ precision at the projected coordinate for interpolation in the reference frame. In order to maintain geometric continuity in the projected frame, we employed sphere padding [12] in the reference frame. This helps in improved prediction along the frame edges. The step sizes $\Delta\phi$ and $\Delta\theta$ in ERP were chosen to be $\frac{2\pi}{W}$ and $\frac{\pi}{H}$, respectively. We measured the distortion in terms of end-to-end weighted spherical PSNR as recommended in [10]. Average bit rate reduction was evaluated using the Bjntegaard model [13]. Bit rate savings over HEVC are tabulated in Table I, and show performance for the Y, U and V components, where encoding is in conjunction with ERP over several sequences. It is evident that the rotational motion compensation approach provides significant overall bit rate reduction of over 11%, on the average, over HEVC for the Y-component. Fig. 3 and 4 show a significant improvement in the visual quality for example frames from the "chairlift" and "driving in country" sequences with the proposed motion model as compared to HEVC based encoding at the same bit rate.

V. Conclusions

The objective of this follow-up paper was to implement and test the rotational motion compensation approach of [1] in a full codec setting that includes bi-directional prediction, multi-reference prediction and hierarchical motion vector refinement, thereby enabling the motion model to cut across all the profiles

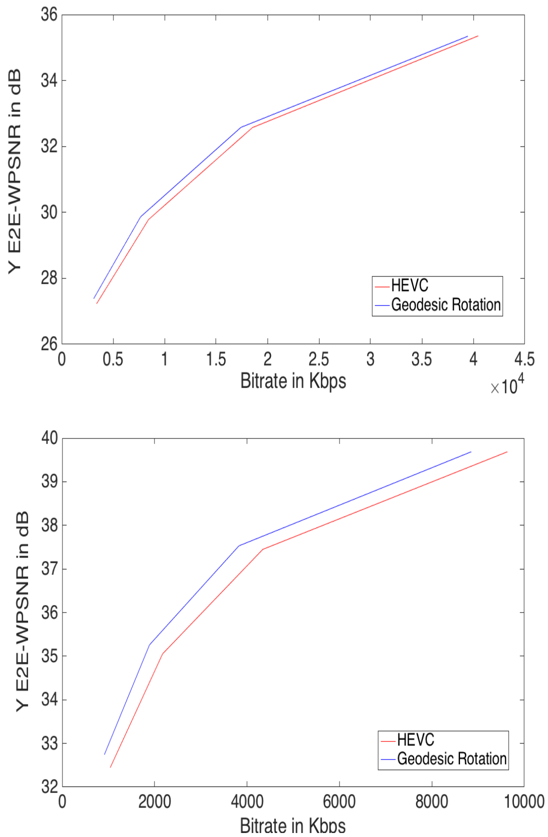


Fig. 2. RD curves for bicyclist (top) and glacier (bottom) sequences

TABLE I
BITRATE SAVINGS IN % OVER HEVC (ERP, RA PROFILE, CTC SEQUENCES)

Sequence	Y	U	V
bicyclist	-12.4	-3.9	-10.5
chair	-15.9	-10.6	-10.3
glacier	-17.9	-6.3	-7.6
brancastle	-6.3	-4.2	-6.2
driving in country	-11.4	-5.2	-5.1
kiteflite	-5.37	-0.4	-4.4
average	-11.5	-5.1	-7.4

in HEVC. The implemented rotation motion model preserves object shape and size, which was observed to be critical for effective bi-directional prediction and multi-reference prediction. In addition, a location-invariant motion search pattern was implemented in direct correspondence with the hierarchical motion vector refinement in HEVC. Substantial gains provide ample evidence for the effectiveness of the rotational motion model.

Fig. 3. Ninth frame of the chairlift sequence: visual comparison between HEVC based encoding (top) and geodesic motion model (bottom)

Fig. 4. Eighth frame of the driving-in-country sequence: visual comparison between HEVC based encoding (top) and geodesic motion model (bottom)

REFERENCES

[1] B. Vishwanath, T. Nanjundaswamy, and K. Rose, "Rotational motion model for temporal prediction in 360 video coding," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2017.

[2] J. P. Snyder, *Flattening the earth: two thousand years of map projections*, University of Chicago Press, 1997.

[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[4] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[5] M. Narroschke and R. Swoboda, "Extending HEVC by an affine motion model," in *Picture Coding Symposium (PCS)*, 2013, pp. 321–324.

[6] L. Li, Z. Li, M. Budagavi, and H. Li, "Projection based advanced motion model for cubic mapping for 360-degree video," *arXiv preprint arXiv:1702.06277*, 2017.

[7] O. Rodrigues, "Des lois géométriques qui regissent les déplacements d'un systéme solide dans l'espace et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire," *Journal de Mathématiques Pures et Appliquées*, vol. 5, pp. 380–440, 1840.

[8] "High efficiency video coding test model, HM-16.15," *https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/ tags/*, 2016.

[9] Y. Ye, E. Alshina, and J. Boyce, "Algorithm descriptions of projection format conversion and video quality metrics in 360lib," *JVET-F1003*, April 2017.

[10] J. Boyce, E. Alshina, A. Abbas, and Y. Ye, "Jvet common test conditions and evaluation procedures for 360 video," *JVET-F1030,*, April 2017.

[11] K. Ugur, A. Alshin, E. Alshina, F. Bossen, W Han, J Park, and J Lainema, "Motion compensated prediction and interpolation filter design in h. 265/hevc," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 946–956, 2013.

[12] Y. He, Y. Ye, P. Hanhart, and X. Xiu, "AHG8: Geometry padding for 360 video coding," *Document JVET-D0075*, 2016.

[13] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4 April*, 2001.